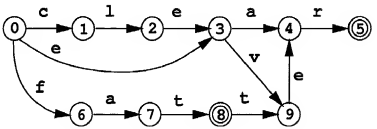
0
1
2
3
4
5
6
7
8
9
c
l
e
a
r
e
v
e
f
a
t
t

FIG. 1

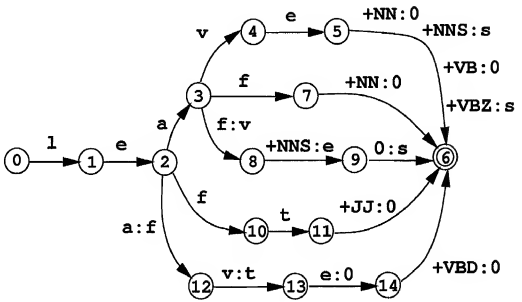
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
l
e
a
v
e
+NN:0
+NNS:s
+VB:0
f
+NN:0
+VBZ:s
f:v
+NNS:e
0:s
f
t
+JJ:0
a:f
v:t
e:0
+VBD:0

FIG. 2

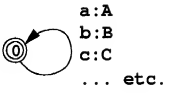

FIG. 3

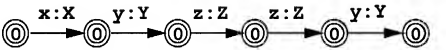

FIG. 4

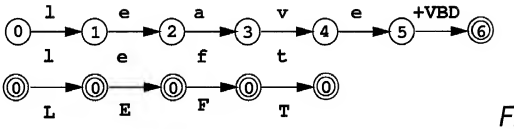

FIG. 5

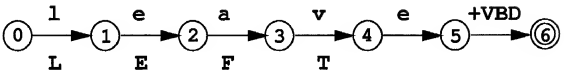

FIG. 6

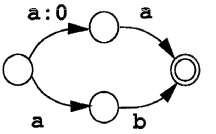

FIG. 7

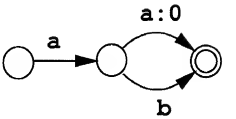

FIG. 8

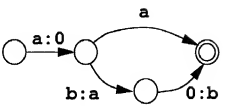

FIG. 9

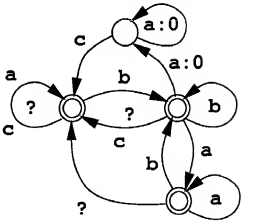

FIG. 10

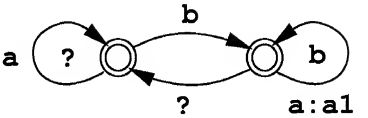

FIG. 11

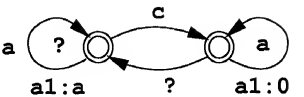

FIG. 12

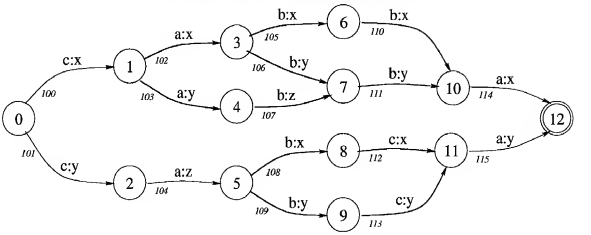
0
1
2
3
4
5
6
7
8
9
10
11
12
c:x
100
a:x
102
a:y
103
b:x
105
b:y
106
b:z
107
b:x
110
b:y
111
a:x
114
c:x
112
a:y
115
c:y
113
b:x
108
b:y
109
c:y
101
a:z
104

FIG. 13

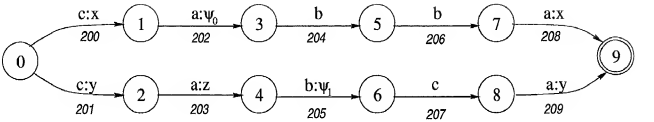
0
1
2
3
4
5
6
7
8
9
c:x
200
a:ψ0
202
b
204
b
206
a:x
208
c:y
201
a:z
203
b:ψ1
205
c
207
a:y
209

FIG. 14

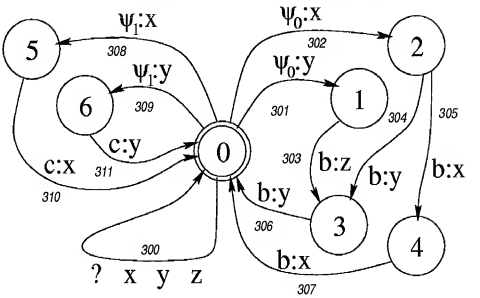
5
6
0
1
2
3
4
ψ1:x
308
ψ1:y
309
c:y
c:x
311
310
300
? x y z
ψ0:x
302
ψ0:y
301
303
b:z
b:y
306
b:y
304
305
b:x
b:x
307

FIG. 15

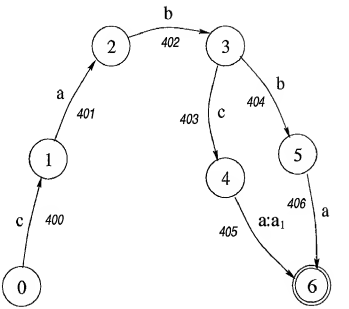
0
1
2
3
4
5
6
c
400
a
401
b
402
403
c
404
b
405
a:a1
406
a

FIG. 16

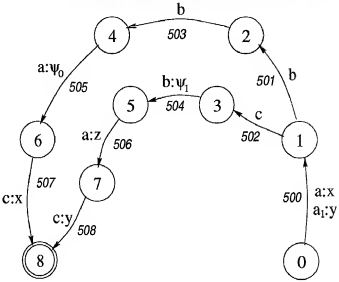
b
503
4
2
a:ψ0
505
b:ψ1
504
5
3
501
b
c
502
1
6
a:z
506
7
c:x
507
c:y
508
8
500
a:x
a1:y
0

FIG. 17

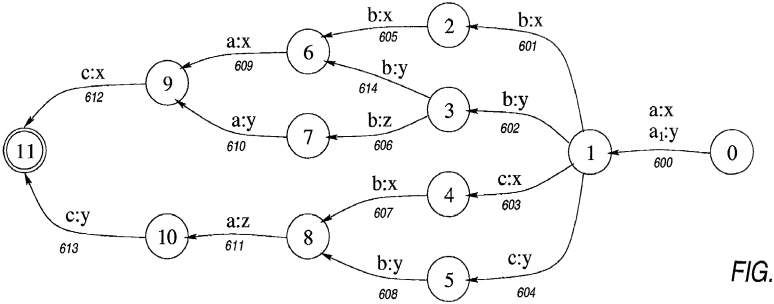
b:x
605
2
b:x
601
a:x
609
6
c:x
612
9
b:y
614
3
b:y
602
a:x
a1:y
600
0
11
a:y
610
7
b:z
606
1
b:x
607
4
c:x
603
c:y
613
10
a:z
611
8
b:y
608
5
c:y
604

FIG. 18

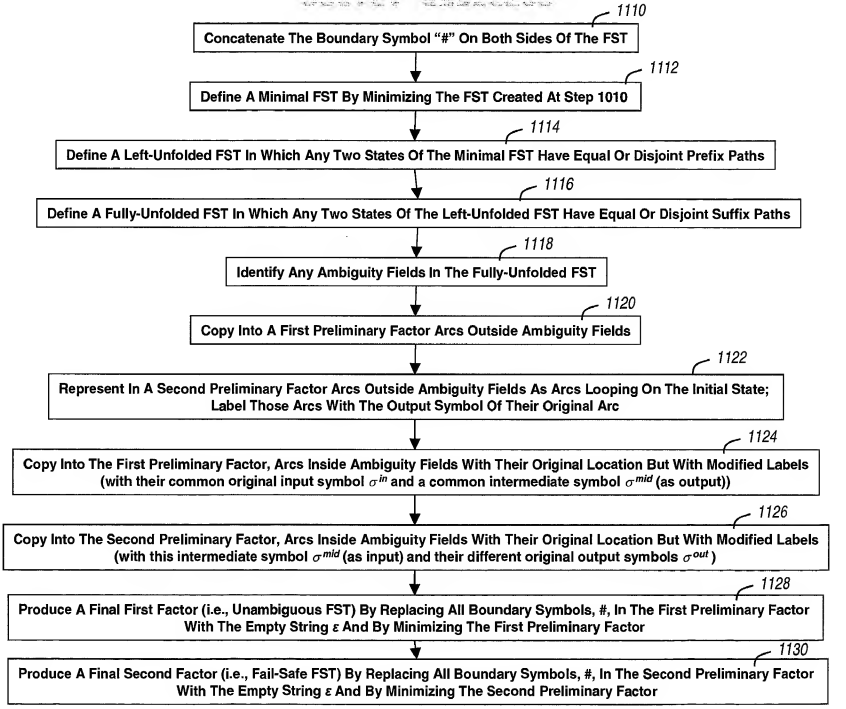
1110
Concatenate The Boundary Symbol "#" On Both Sides Of The FST
1112
Define A Minimal FST By Minimizing The FST Created At Step 1010
1114
Define A Left-Unfolded FST In Which Any Two States Of The Minimal FST Have Equal Or Disjoint Prefix Paths
1116
Define A Fully-Unfolded FST In Which Any Two States Of The Left-Unfolded FST Have Equal Or Disjoint Suffix Paths
1118
Identify Any Ambiguity Fields In The Fully-Unfolded FST
1120
Copy Into A First Preliminary Factor Arcs Outside Ambiguity Fields
1122
Represent In A Second Preliminary Factor Arcs Outside Ambiguity Fields As Arcs Looping On The Initial State; Label Those Arcs With The Output Symbol Of Their Original Arc
1124
Copy Into The First Preliminary Factor, Arcs Inside Ambiguity Fields With Their Original Location But With Modified Labels (with their common original input symbol σ^{in} and a common intermediate symbol σ^{mid} (as output))
1126
Copy Into The Second Preliminary Factor, Arcs Inside Ambiguity Fields With Their Original Location But With Modified Labels (with this intermediate symbol σ^{mid} (as input) and their different original output symbols σ^{out})
1128
Produce A Final First Factor (i.e., Unambiguous FST) By Replacing All Boundary Symbols, #, In The First Preliminary Factor With The Empty String ε And By Minimizing The First Preliminary Factor
1130
Produce A Final Second Factor (i.e., Fail-Safe FST) By Replacing All Boundary Symbols, #, In The Second Preliminary Factor With The Empty String ε And By Minimizing The Second Preliminary Factor

FIG. 19

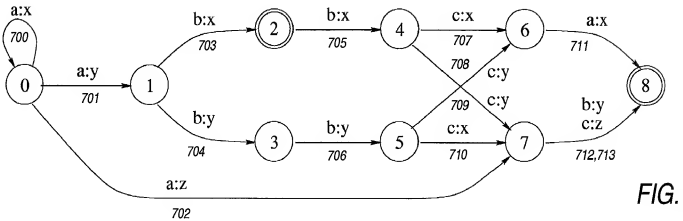
a:x
700
0
a:y
701
1
b:x
703
2
b:x
705
4
c:x
707
6
a:x
711
8
b:y
704
3
b:y
706
5
c:x
710
7
712,713
c:y
708
c:y
709
b:y
c:z
a:z
702

FIG. 20

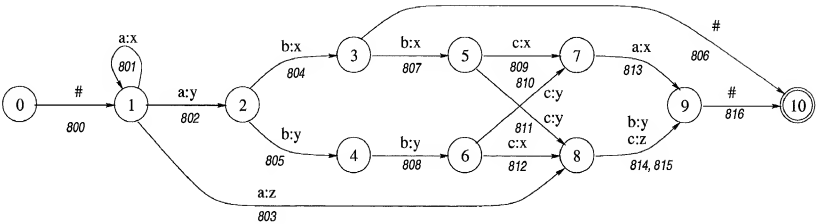
0
#
800
a:x
801
1
a:y
802
2
b:x
804
3
b:x
807
5
c:x
809
7
a:x
813
9
#
816
10
#
806
c:y
810
c:y
811
b:y
805
4
b:y
808
6
c:x
812
8
b:y
c:z
814, 815
a:z
803

FIG. 21

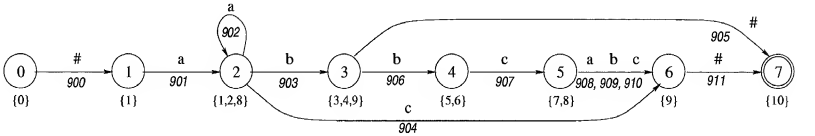
0
1
2
3
4
5
6
7
#
a
b
c
900
901
902
903
904
905
906
907
908, 909, 910
911
{0}
{1}
{1,2,8}
{3,4,9}
{5,6}
{7,8}
{9}
{10}

FIG. 22

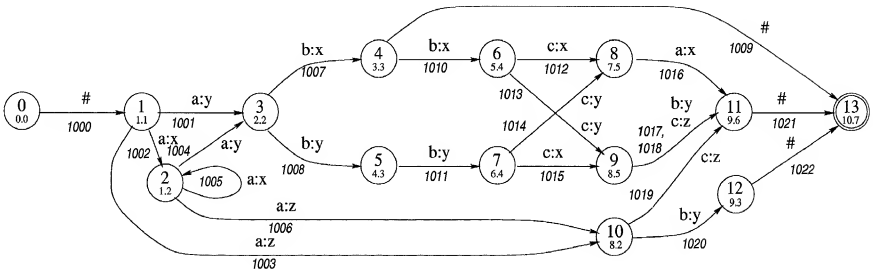
0
0.0
1
1.1
2
1.2
3
2.2
4
3.3
5
4.3
6
5.4
7
6.4
8
7.5
9
8.5
10
8.2
11
9.6
12
9.3
13
10.7
#
a:y
a:x
b:x
b:y
c:x
c:y
c:z
a:z
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017,
1018
1019
1020
1021
1022

FIG. 23

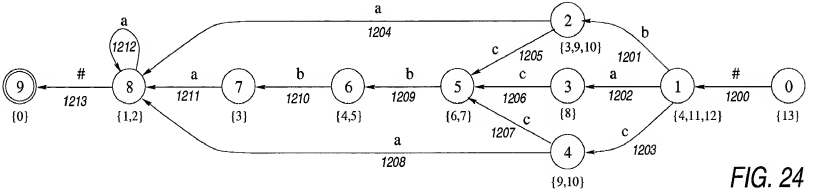
9
{0}
#
1213
8
{1,2}
a
1212
a
1211
7
{3}
b
1210
6
{4,5}
b
1209
5
{6,7}
c
1205
2
{3,9,10}
a
1204
c
1206
3
{8}
b
1201
a
1202
1
{4,11,12}
#
1200
0
{13}
c
1207
4
{9,10}
a
1208
c
1203

FIG. 24

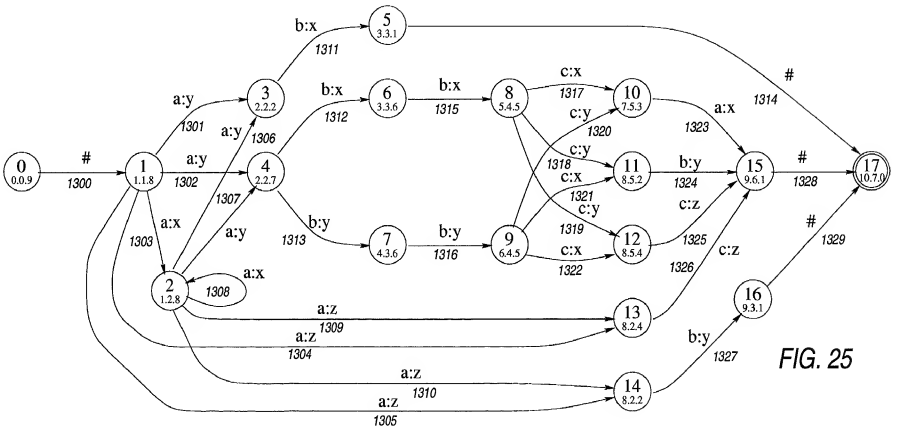
0
0.0.9
#
1300
1
1.1.8
2
1.2.8
3
2.2.2
4
2.2.7
5
3.3.1
6
3.3.6
7
4.3.6
8
5.4.5
9
6.4.5
10
7.5.3
11
8.5.2
12
8.5.4
13
8.2.4
14
8.2.2
15
9.6.1
16
9.3.1
17
10.7.0
a:y
1301
a:y
1302
a:y
1306
a:x
1303
a:y
1307
a:x
1308
b:x
1311
b:x
1312
b:y
1313
b:x
1315
b:y
1316
c:x
1317
c:y
1320
c:y
1318
c:x
1321
c:y
1319
c:x
1322
a:x
1323
b:y
1324
c:z
1325
c:z
1326
b:y
1327
#
1314
#
1328
#
1329
a:z
1309
a:z
1304
a:z
1310
a:z
1305

FIG. 25

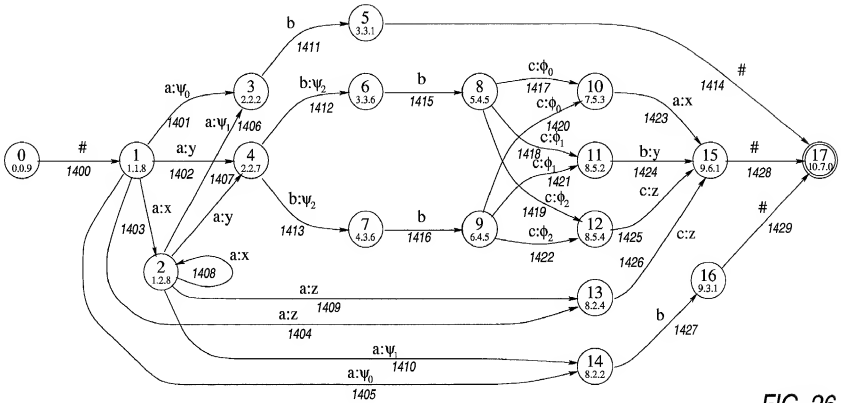
0
0.0.9
1
1.1.8
2
1.2.8
3
2.2.2
4
2.2.7
5
3.3.1
6
3.3.6
7
4.3.6
8
5.4.5
9
6.4.5
10
7.5.3
11
8.5.2
12
8.5.4
13
8.2.4
14
8.2.2
15
9.6.1
16
9.3.1
17
10.7.9
#
1400
a:y
1402
a:x
1403
a:y
a:x
1408
a:z
1409
a:z
1404
a:ψ1
1410
a:ψ0
1405
a:ψ0
1401
a:ψ1
1406
b
1411
b:ψ2
1412
b:ψ2
1413
b
1415
b
1416
c:ϕ0
1417
c:ϕ0
1420
c:ϕ1
1418
c:ϕ1
1421
c:ϕ2
1419
c:ϕ2
1422
a:x
1423
b:y
1424
c:z
1425
c:z
1426
b
1427
#
1414
#
1428
#
1429

FIG. 26

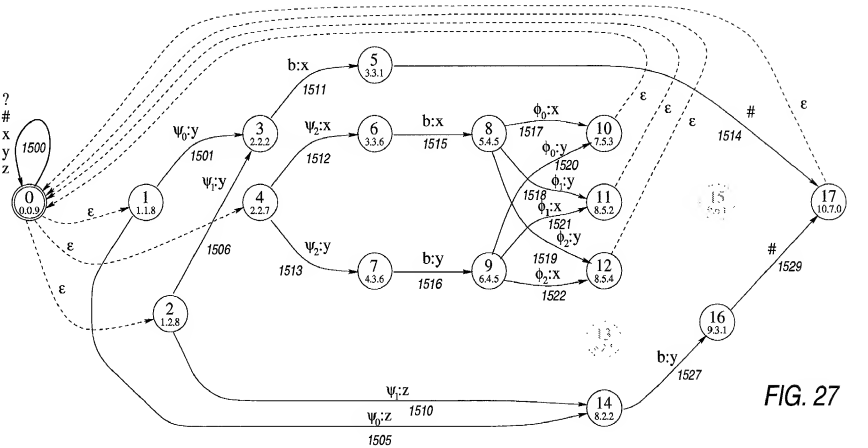
?
#
x
y
z
1500
0
0.0.9
ε
1
1.1.8
ε
ε
2
1.2.8
1506
ψ0:y
1501
3
2.2.2
ψ1:y
4
2.2.7
b:x
1511
5
3.3.1
ψ2:x
1512
6
3.3.6
ψ2:y
1513
7
4.3.6
b:x
1515
b:y
1516
8
5.4.5
9
6.4.5
φ0:x
1517
φ0:y
1520
φ1:y
1518
φ1:x
1521
φ2:y
1519
φ2:x
1522
10
7.5.3
11
8.5.2
12
8.5.4
ε
ε
ε
#
1514
ε
17
10.7.0
#
1529
16
9.3.1
b:y
1527
ψ1:z
ψ0:z
1510
1505
14
8.2.2

FIG. 27

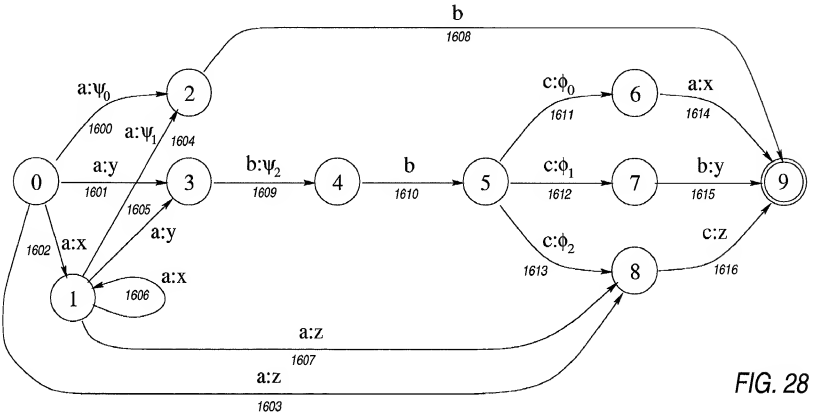
b
1608
a:ψ0
2
c:ϕ0
6
a:x
1600
a:ψ1
1604
1611
1614
0
a:y
1601
3
b:ψ2
1609
4
b
1610
5
c:ϕ1
1612
7
b:y
1615
9
1605
a:y
a:x
1602
c:ϕ2
1613
8
c:z
1616
a:x
1606
1
a:z
1607
a:z
1603

FIG. 28

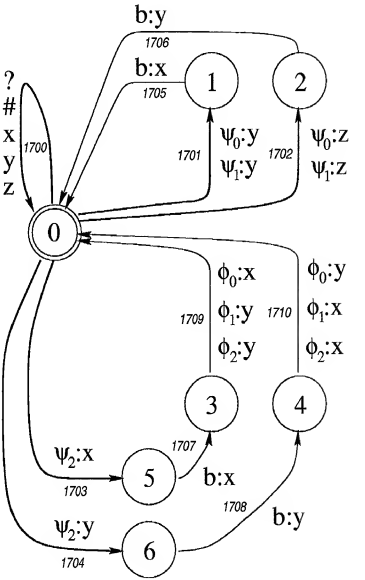
b:y
1706
b:x
1705
?
#
x
y
z
1700
1
2
ψ_0:y
ψ_1:y
1701
ψ_0:z
ψ_1:z
1702
0
ϕ_0:x
ϕ_1:y
ϕ_2:y
1709
ϕ_0:y
ϕ_1:x
ϕ_2:x
1710
3
4
1707
ψ_2:x
1703
5
b:x
1708
b:y
ψ_2:y
1704
6

FIG. 29

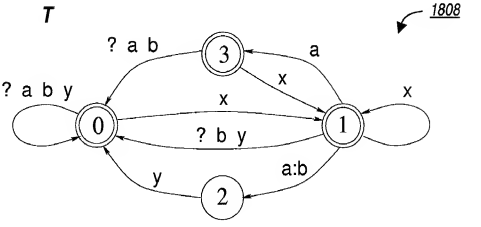
T
1808
? a b
3
a
x
? a b y
x
x
0
1
? b y
a:b
y
2

FIG. 30

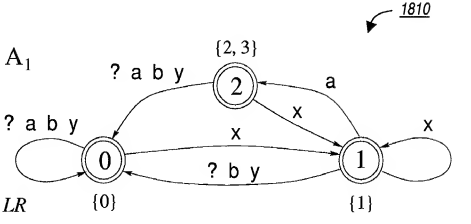
1810
A1
{2, 3}
? a b y
2
a
x
? a b y
x
x
0
1
? b y
LR
{0}
{1}

FIG. 31

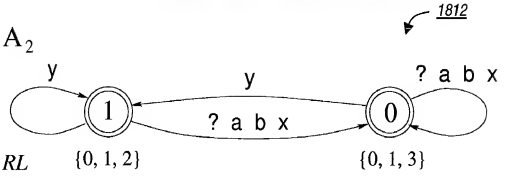
1812
A2
y
y
? a b x
1
? a b x
0
RL
{0, 1, 2}
{0, 1, 3}

FIG. 32

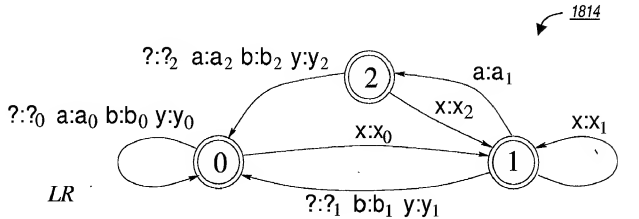
1814
?:?2 a:a2 b:b2 y:y2
2
a:a1
?:?0 a:a0 b:b0 y:y0
x:x2
x:x1
T1
x:x0
0
1
LR
?:?1 b:b1 y:y1

FIG. 33

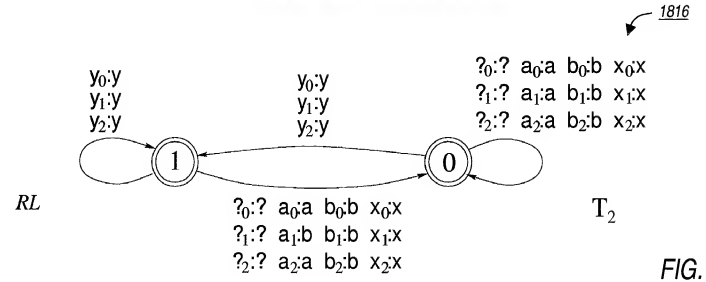
1816
y0:y
y1:y
y2:y
y0:y
y1:y
y2:y
?0:? a0:a b0:b x0:x
?1:? a1:a b1:b x1:x
?2:? a2:a b2:b x2:x
1
0
RL
?0:? a0:a b0:b x0:x
?1:? a1:b b1:b x1:x
?2:? a2:a b2:b x2:x
T2

FIG. 34

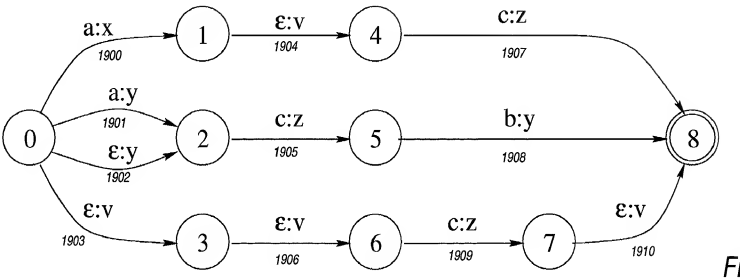
a:x
1900
1
ε:v
1904
4
c:z
1907
a:y
1901
0
2
c:z
1905
5
b:y
1908
8
ε:y
1902
ε:v
1903
3
ε:v
1906
6
c:z
1909
7
ε:v
1910

FIG. 35

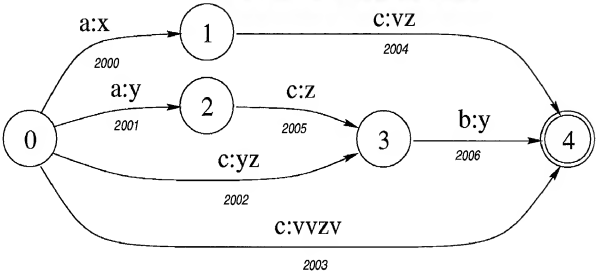

FIG. 36

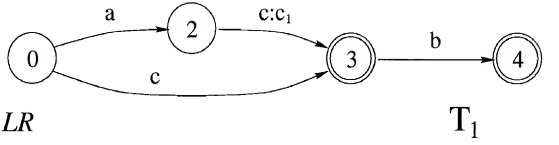

FIG. 37

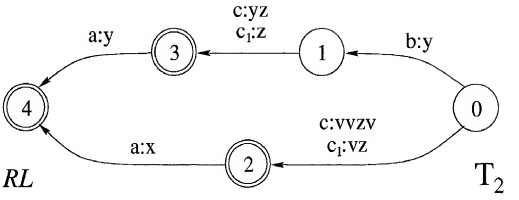

FIG. 38

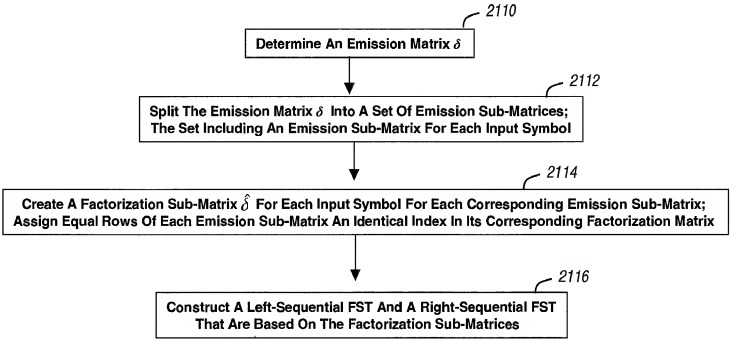
2110
Determine An Emission Matrix δ
2112
Split The Emission Matrix δ Into A Set Of Emission Sub-Matrices; The Set Including An Emission Sub-Matrix For Each Input Symbol
2114
Create A Factorization Sub-Matrix $\hat{\delta}$ For Each Input Symbol For Each Corresponding Emission Sub-Matrix; Assign Equal Rows Of Each Emission Sub-Matrix An Identical Index In Its Corresponding Factorization Matrix
2116
Construct A Left-Sequential FST And A Right-Sequential FST That Are Based On The Factorization Sub-Matrices

FIG. 39

09973952-1E1800

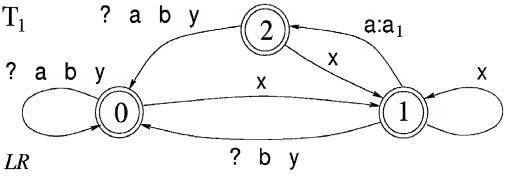

FIG. 40

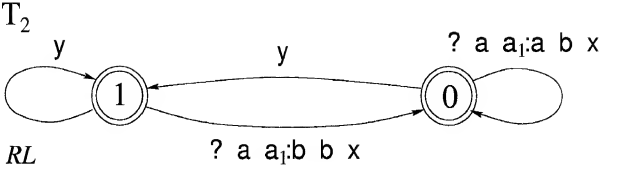

FIG. 41

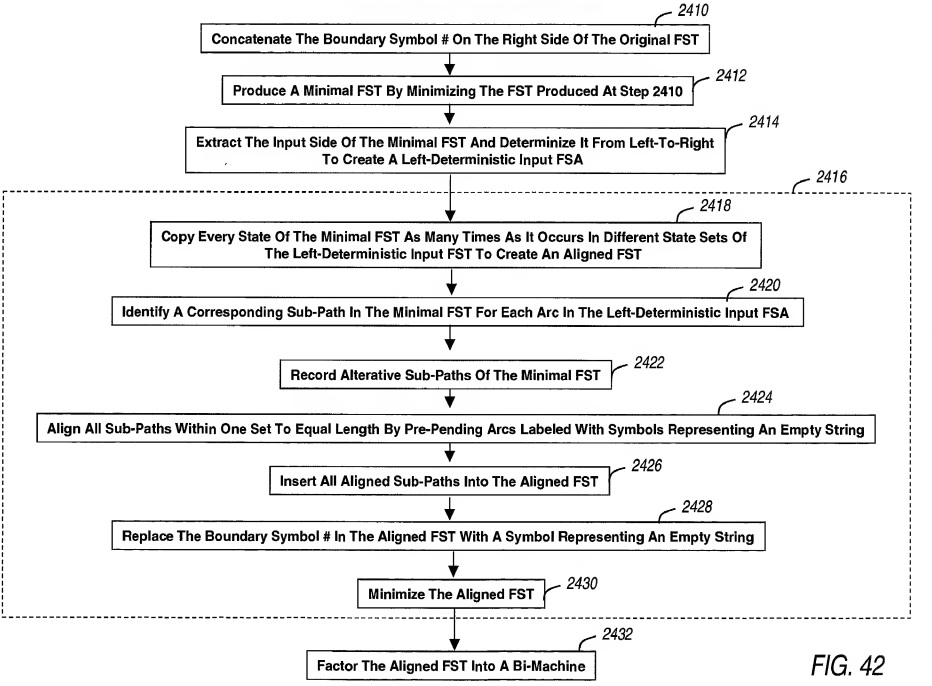
2410
Concatenate The Boundary Symbol # On The Right Side Of The Original FST
2412
Produce A Minimal FST By Minimizing The FST Produced At Step 2410
2414
Extract The Input Side Of The Minimal FST And Determinize It From Left-To-Right To Create A Left-Deterministic Input FSA
2416
2418
Copy Every State Of The Minimal FST As Many Times As It Occurs In Different State Sets Of The Left-Deterministic Input FST To Create An Aligned FST
2420
Identify A Corresponding Sub-Path In The Minimal FST For Each Arc In The Left-Deterministic Input FSA
2422
Record Alterative Sub-Paths Of The Minimal FST
2424
Align All Sub-Paths Within One Set To Equal Length By Pre-Pending Arcs Labeled With Symbols Representing An Empty String
2426
Insert All Aligned Sub-Paths Into The Aligned FST
2428
Replace The Boundary Symbol # In The Aligned FST With A Symbol Representing An Empty String
2430
Minimize The Aligned FST
2432
Factor The Aligned FST Into A Bi-Machine

FIG. 42

UNDER SEAL

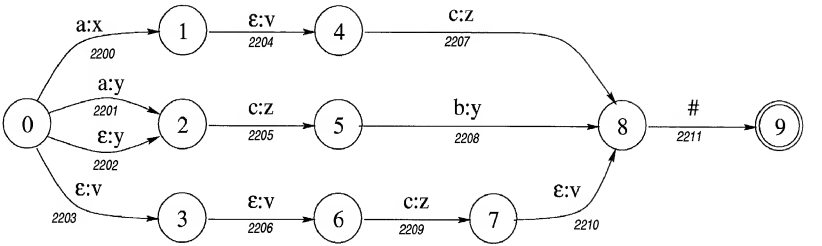

FIG. 43

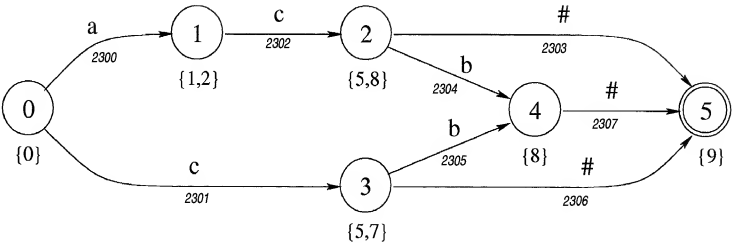

FIG. 44

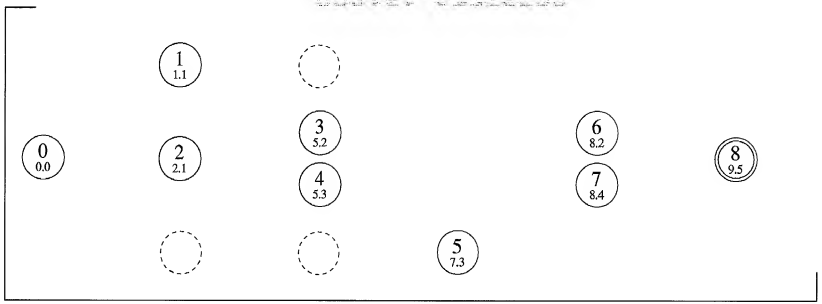
0
0.0
1
1.1
2
2.1
3
5.2
4
5.3
5
7.3
6
8.2
7
8.4
8
9.5

FIG. 45

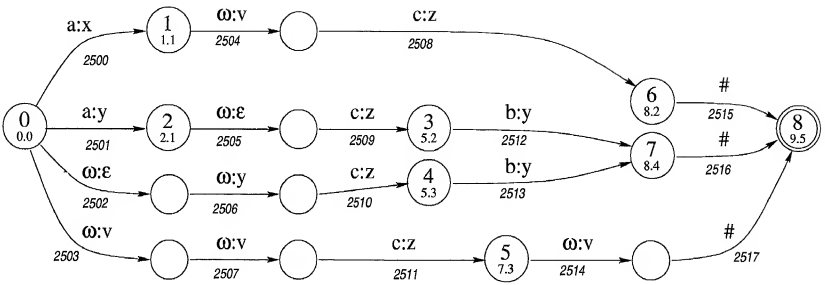
0
0.0
1
1.1
2
2.1
3
5.2
4
5.3
5
7.3
6
8.2
7
8.4
8
9.5
a:x
2500
a:y
2501
ω:ε
2502
ω:v
2503
ω:v
2504
ω:ε
2505
ω:y
2506
ω:v
2507
c:z
2508
c:z
2509
c:z
2510
c:z
2511
b:y
2512
b:y
2513
ω:v
2514
#
2515
#
2516
#
2517

FIG. 46

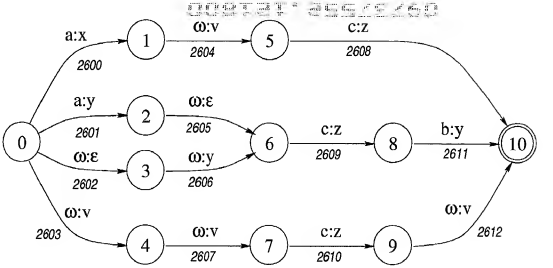

FIG. 47

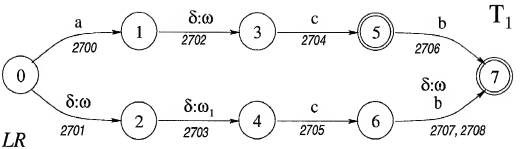

FIG. 48

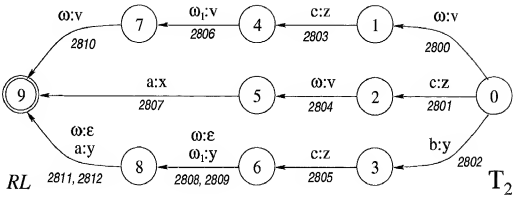

FIG. 49

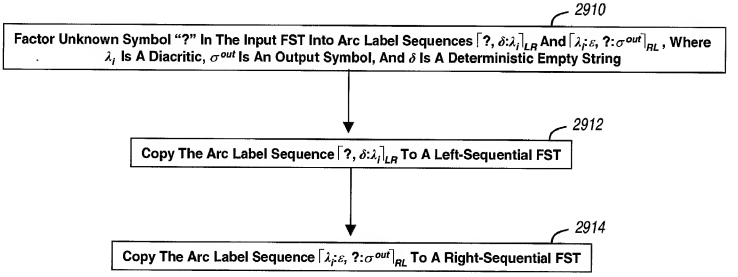
2910
Factor Unknown Symbol "?" In The Input FST Into Arc Label Sequences $\lceil ?, \delta{:}\lambda_i \rceil_{LR}$ And $\lceil \lambda_i{:}\varepsilon, ?{:}\sigma^{out} \rceil_{RL}$, Where λ_i Is A Diacritic, σ^{out} Is An Output Symbol, And δ Is A Deterministic Empty String
2912
Copy The Arc Label Sequence $\lceil ?, \delta{:}\lambda_i \rceil_{LR}$ To A Left-Sequential FST
2914
Copy The Arc Label Sequence $\lceil \lambda_i{:}\varepsilon, ?{:}\sigma^{out} \rceil_{RL}$ To A Right-Sequential FST

FIG. 50

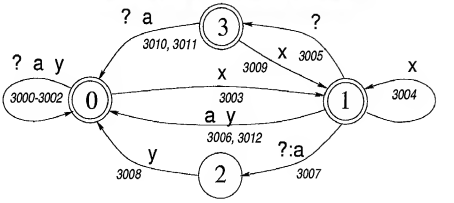

FIG. 51

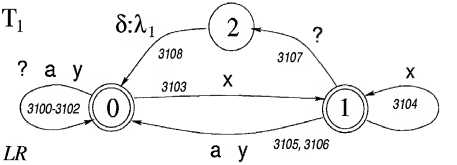

FIG. 52

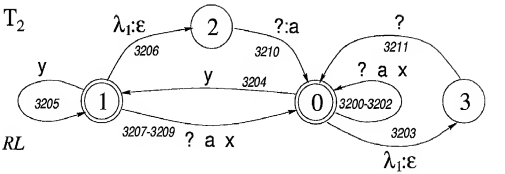

FIG. 53

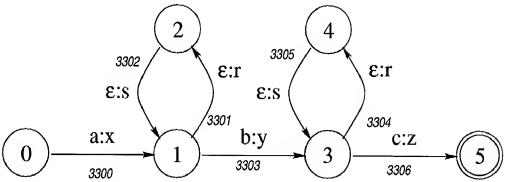
2
4
3302
3305
ε:r
ε:r
ε:s
ε:s
3301
3304
0
a:x
1
b:y
3
c:z
5
3300
3303
3306

FIG. 54

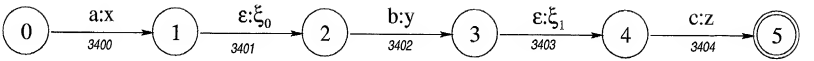
0
a:x
3400
1
ε:ξ0
3401
2
b:y
3402
3
ε:ξ1
3403
4
c:z
3404
5

FIG. 55

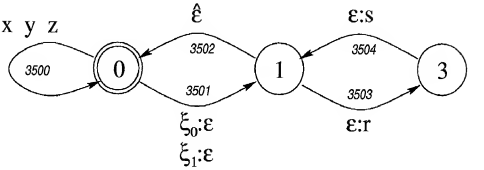
x y z
3500
0
ε̂
3502
3501
ξ0:ε
ξ1:ε
1
ε:s
3504
3503
ε:r
3

FIG. 56

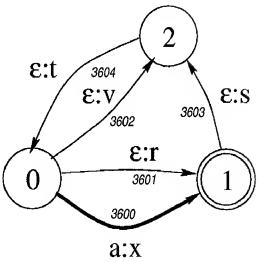
2
ε:t
3604
ε:v
3602
ε:s
3603
ε:r
0
3601
1
3600
a:x

FIG. 57

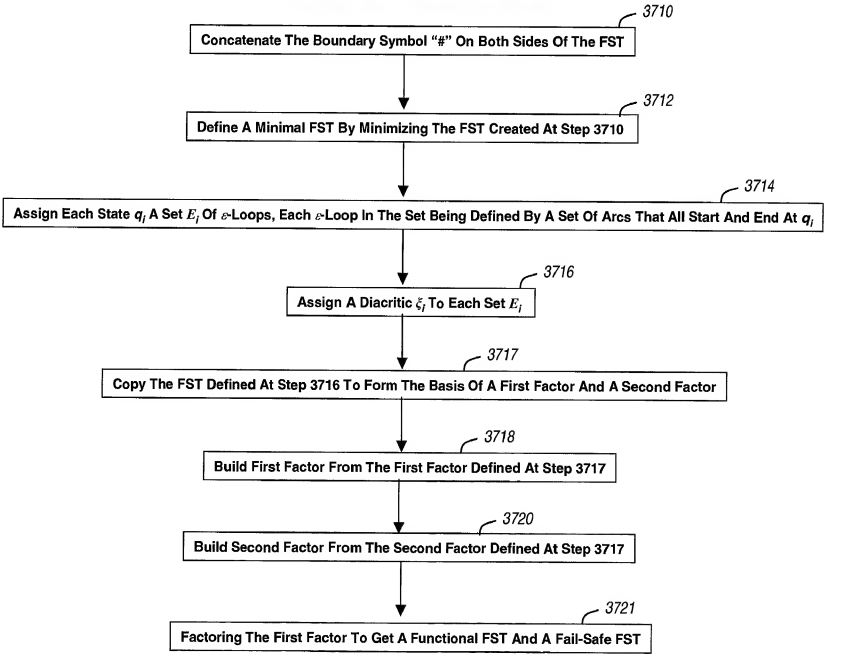
3710
Concatenate The Boundary Symbol "#" On Both Sides Of The FST
3712
Define A Minimal FST By Minimizing The FST Created At Step 3710
3714
Assign Each State q_i A Set E_i Of ε-Loops, Each ε-Loop In The Set Being Defined By A Set Of Arcs That All Start And End At q_i
3716
Assign A Diacritic ξ_i To Each Set E_i
3717
Copy The FST Defined At Step 3716 To Form The Basis Of A First Factor And A Second Factor
3718
Build First Factor From The First Factor Defined At Step 3717
3720
Build Second Factor From The Second Factor Defined At Step 3717
3721
Factoring The First Factor To Get A Functional FST And A Fail-Safe FST

FIG. 58

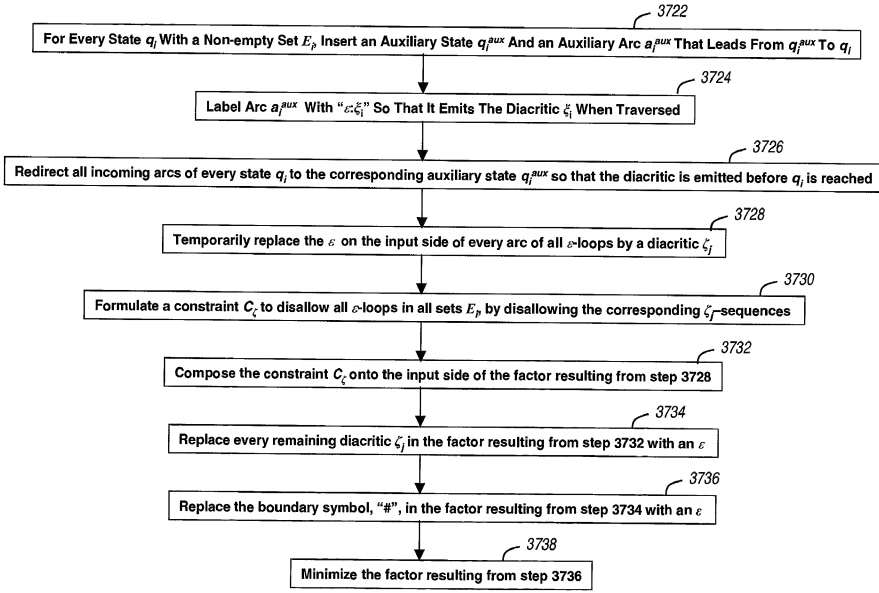
3722
For Every State q_i With a Non-empty Set E_i, Insert an Auxiliary State q_i^{aux} And an Auxiliary Arc a_i^{aux} That Leads From q_i^{aux} To q_i
3724
Label Arc a_i^{aux} With "$\varepsilon:\xi_i$" So That It Emits The Diacritic ξ_i When Traversed
3726
Redirect all incoming arcs of every state q_i to the corresponding auxiliary state q_i^{aux} so that the diacritic is emitted before q_i is reached
3728
Temporarily replace the ε on the input side of every arc of all ε-loops by a diacritic ζ_j
3730
Formulate a constraint C_ζ to disallow all ε-loops in all sets E_i by disallowing the corresponding ζ_j-sequences
3732
Compose the constraint C_ζ onto the input side of the factor resulting from step 3728
3734
Replace every remaining diacritic ζ_j in the factor resulting from step 3732 with an ε
3736
Replace the boundary symbol, "#", in the factor resulting from step 3734 with an ε
3738
Minimize the factor resulting from step 3736

FIG. 59

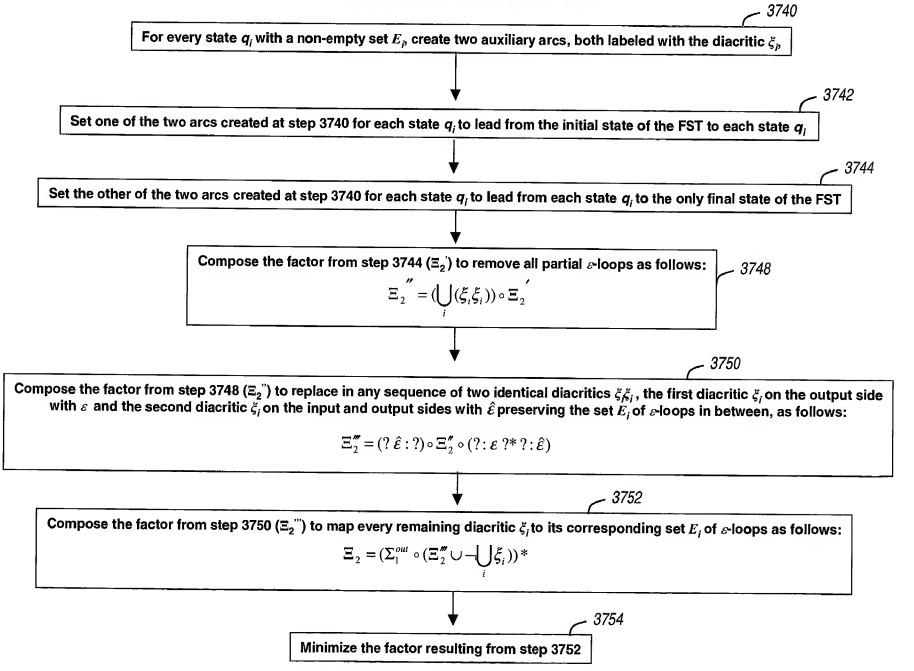
3740
For every state q_i with a non-empty set E_p create two auxiliary arcs, both labeled with the diacritic ξ_p
3742
Set one of the two arcs created at step 3740 for each state q_i to lead from the initial state of the FST to each state q_i
3744
Set the other of the two arcs created at step 3740 for each state q_i to lead from each state q_i to the only final state of the FST
Compose the factor from step 3744 (Ξ_2') to remove all partial ε-loops as follows:
$\Xi_2'' = (\bigcup_i (\xi_i \xi_i)) \circ \Xi_2'$
3748
3750
Compose the factor from step 3748 (Ξ_2'') to replace in any sequence of two identical diacritics $\xi_i \xi_i$, the first diacritic ξ_i on the output side with ε and the second diacritic ξ_i on the input and output sides with $\hat{\varepsilon}$ preserving the set E_i of ε-loops in between, as follows:
$\Xi_2''' = (?\ \hat{\varepsilon} : ?) \circ \Xi_2'' \circ (? : \varepsilon\ ?^* ? : \hat{\varepsilon})$
3752
Compose the factor from step 3750 (Ξ_2''') to map every remaining diacritic ξ_i to its corresponding set E_i of ε-loops as follows:
$\Xi_2 = (\Sigma_1^{out} \circ (\Xi_2''' \cup \neg(\bigcup_i \xi_i))^*$
3754
Minimize the factor resulting from step 3752

FIG. 60

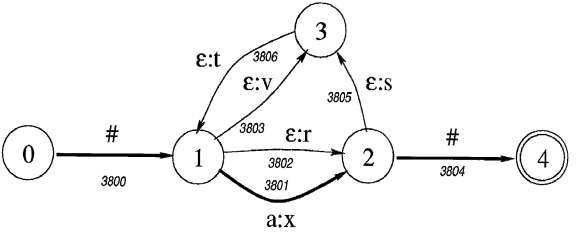
3
ε:t
3806
ε:v
ε:s
3805
0
#
3800
1
3803
ε:r
3802
3801
2
#
3804
4
a:x

FIG. 61

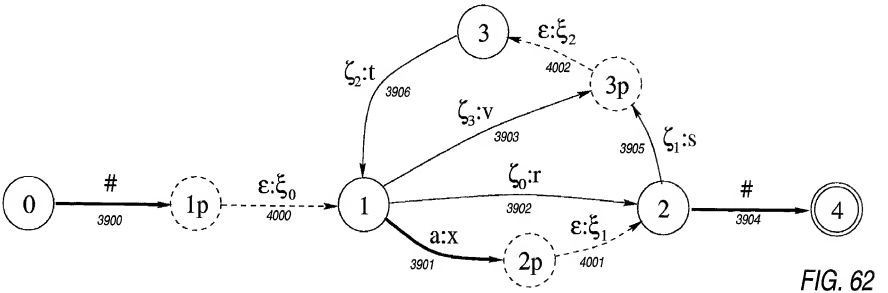
3
ε:ξ2
4002
3p
ζ2:t
3906
ζ3:v
3903
3905
ζ1:s
0
#
3900
1p
ε:ξ0
4000
1
ζ0:r
3902
2
#
3904
4
a:x
3901
2p
ε:ξ1
4001

FIG. 62

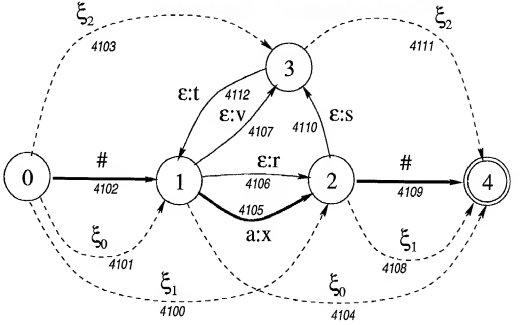
ξ2
4103
ξ2
4111
3
ε:t
4112
ε:v
4107
4110
ε:s
0
#
4102
1
ε:r
4106
2
#
4109
4
4105
a:x
ξ0
4101
ξ1
4108
ξ1
4100
ξ0
4104

FIG. 63

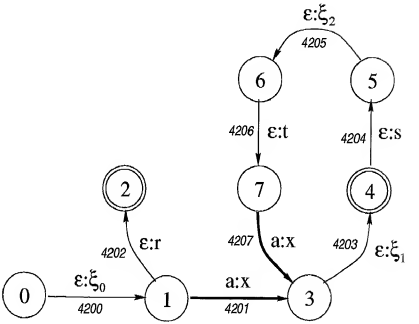
ε:ξ2
4205
6
5
4206
ε:t
4204
ε:s
2
7
4
4202
ε:r
4207
a:x
4203
ε:ξ1
0
ε:ξ0
4200
1
a:x
4201
3

FIG. 64

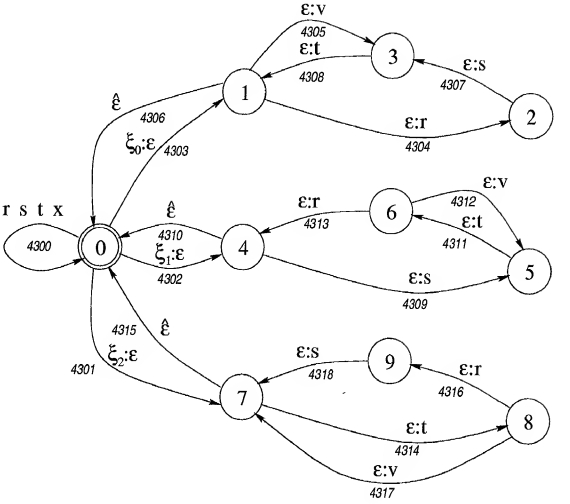
r s t x
4300
0
4301
4302
4303
4304
4305
4306
4307
4308
4309
4310
4311
4312
4313
4314
4315
4316
4317
4318
1
2
3
4
5
6
7
8
9
ε:v
ε:t
ε:s
ε:r
ε̂
ξ0:ε
ξ1:ε
ξ2:ε

FIG. 65

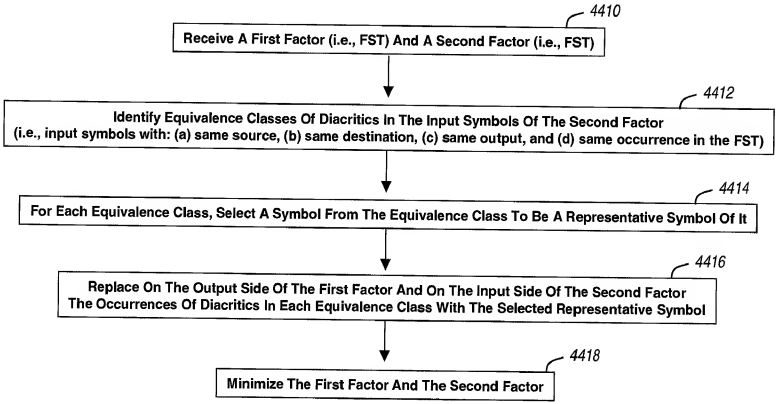
4410
Receive A First Factor (i.e., FST) And A Second Factor (i.e., FST)
4412
Identify Equivalence Classes Of Diacritics In The Input Symbols Of The Second Factor
(i.e., input symbols with: (a) same source, (b) same destination, (c) same output, and (d) same occurrence in the FST)
4414
For Each Equivalence Class, Select A Symbol From The Equivalence Class To Be A Representative Symbol Of It
4416
Replace On The Output Side Of The First Factor And On The Input Side Of The Second Factor
The Occurrences Of Diacritics In Each Equivalence Class With The Selected Representative Symbol
4418
Minimize The First Factor And The Second Factor

FIG. 66

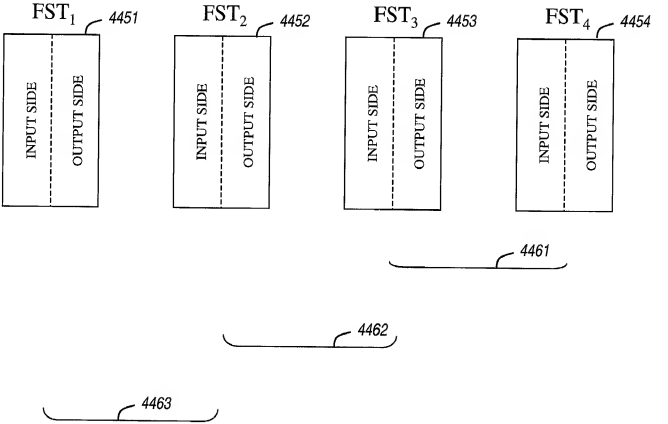
FST1
4451
INPUT SIDE
OUTPUT SIDE
FST2
4452
INPUT SIDE
OUTPUT SIDE
FST3
4453
INPUT SIDE
OUTPUT SIDE
FST4
4454
INPUT SIDE
OUTPUT SIDE
4461
4462
4463

FIG. 67

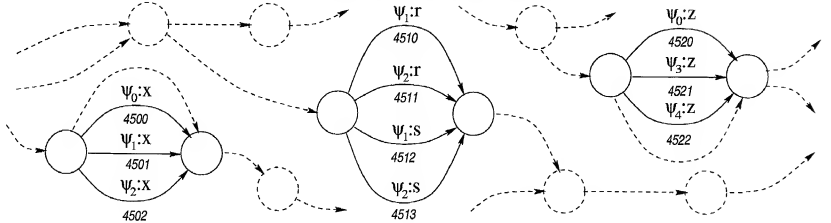
ψ0:x
4500
ψ1:x
4501
ψ2:x
4502
ψ1:r
4510
ψ2:r
4511
ψ1:s
4512
ψ2:s
4513
ψ0:z
4520
ψ3:z
4521
ψ4:z
4522

FIG. 68

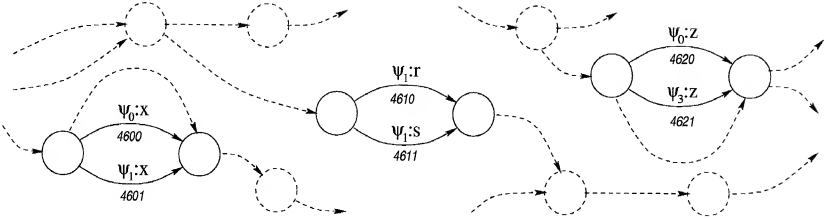
ψ0:x
4600
ψ1:x
4601
ψ1:r
4610
ψ1:s
4611
ψ0:z
4620
ψ3:z
4621

FIG. 69

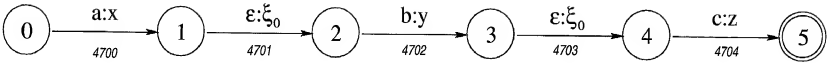
0
a:x
4700
1
ε:ξ0
4701
2
b:y
4702
3
ε:ξ0
4703
4
c:z
4704
5

FIG. 70

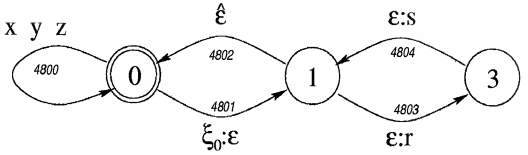
x y z
4800
0
ε̂
4802
4801
ξ0:ε
1
ε:s
4804
4803
ε:r
3

FIG. 71

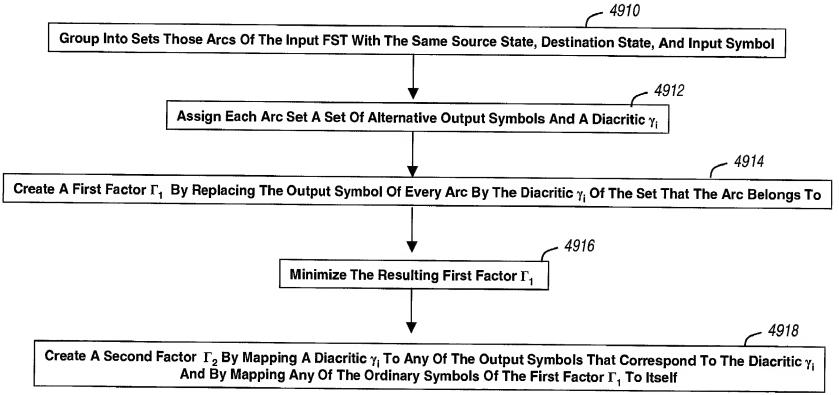
4910
Group Into Sets Those Arcs Of The Input FST With The Same Source State, Destination State, And Input Symbol
4912
Assign Each Arc Set A Set Of Alternative Output Symbols And A Diacritic γ_i
4914
Create A First Factor Γ_1 By Replacing The Output Symbol Of Every Arc By The Diacritic γ_i Of The Set That The Arc Belongs To
4916
Minimize The Resulting First Factor Γ_1
4918
Create A Second Factor Γ_2 By Mapping A Diacritic γ_i To Any Of The Output Symbols That Correspond To The Diacritic γ_i And By Mapping Any Of The Ordinary Symbols Of The First Factor Γ_1 To Itself

FIG. 72

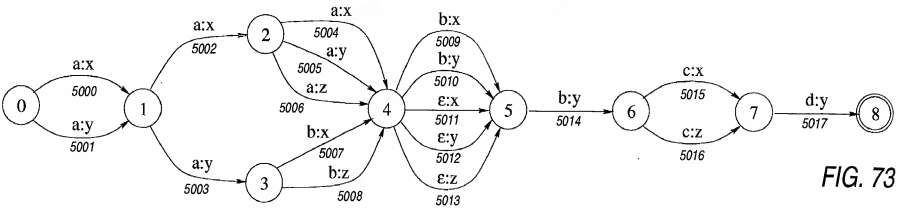

FIG. 73

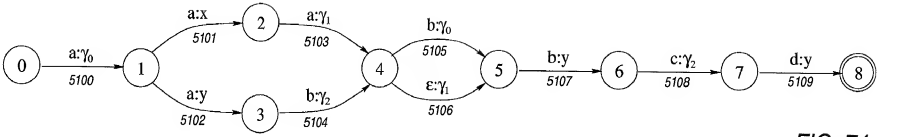

FIG. 74

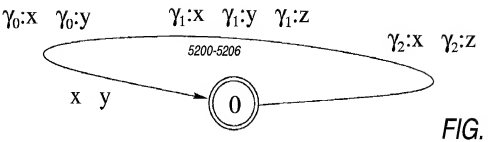

FIG. 75

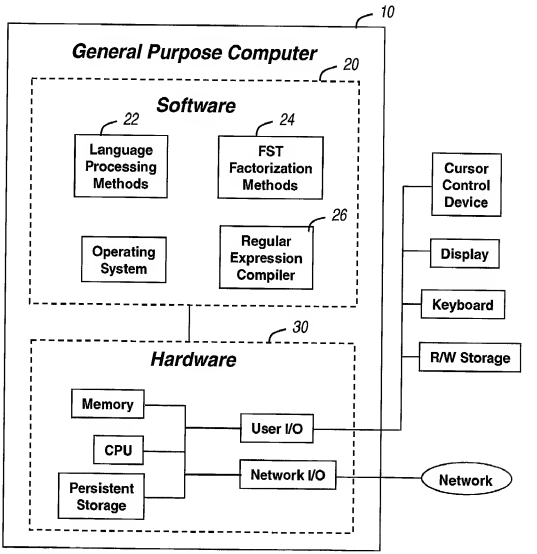
10
General Purpose Computer
20
Software
22
Language Processing Methods
24
FST Factorization Methods
26
Regular Expression Compiler
Operating System
30
Hardware
Memory
CPU
Persistent Storage
User I/O
Network I/O
Cursor Control Device
Display
Keyboard
R/W Storage
Network

FIG. 76